Evaluating Research: How Can You Tell the Good Research from the Bad?

The major criteria used to evaluate scientific research are grouped under the headings "reliability" and "validity." **Reliability** is the degree to which a measure is free of measurement error. It is typically assessed by looking at the consistency of a measure. For example, a reliable weight scale will produce a consistent weight estimate for the same object over time. A reliable measure of playground aggression will produce consistent aggression estimates for the same child across several observers. We will discuss reliability in greater detail in the reading on Measurement. There are several types of **validity**, and each refers to the confidence you can have in making certain conclusions. We will discuss each in greater depth in the sections on validity, but they are presented here to begin familiarizing you with them.

Internal Validity

Internal validity is the confidence you can have in making **cause-and-effect conclusions** from the results of your study. For example, let's say you want to determine whether smoking cigarettes causes lung cancer. The problem is that there are many possible causes of lung cancer – exposure to environmental toxins, genetic predispositions, etc. A study has internal validity to the degree that it is able to rule out these alternative causes and conclude that one factor alone (in this case smoking) produced an effect. We will discuss how this is done in the sections on validity and experimental design.

External Validity

External validity is the confidence you can have in **generalizing** the findings of your study to people, settings, and times not included in your study. Concerns about external validity are generally concerns about the sample used in a study. If you learn that a sample was all-male or all-college-student or all-White, you may have doubts about how well the results would generalize beyond those groups.

Construct Validity

Construct validity is the confidence you can have that the *theoretical* constructs you are attempting to represent in your study (e.g., time pressure, anger) are accurately represented by your *real-world* methods. For example, let's say you are conducting an experiment in which one group is supposed to feel more time pressure than another. This could be done with poor construct validity, for example by telling one group that they should feel rushed, or it could be done with better construct validity, for example by telling students that a test measures their potential for professional success and not giving them enough time to finish it.

Types of Research

Hypothesis-Testing and Exploratory Research

Hypothesis testing, as the name implies, is guided by hypotheses (testable predictions). The use of hypotheses in research improves falsifiability. By firmly stating your expectations before you collect data, you will know when you are incorrect. Thus, erroneous explanations for behavior are more easily identified, and discounted, when hypothesis-testing rather than exploratory research is used. However, the tight focus on a hypothesis can restrict a researcher's perspective, preventing him or her from noticing an important relationship.

Exploratory research aims to provide an objective, detailed description of a phenomenon. For example, researchers working for eBay might be curious about factors unrelated to the product itself that influence the final sale price of the product. They might correlate final sale price with all of the variables related to item listing: whether a photo is included, the starting bid, use of exclamation marks, reputation of seller, etc. Exploratory research cares less about testing a particular hypothesis (e.g., adding a photo increases sales) and more about providing information about a phenomenon. It is more common early in a research program, when little is known about a phenomenon.

Quantitative and Qualitative Research

Quantitative

In research that is **quantitative**, observations are transformed into codes according to some consistent set of rules. For example, the "ambitiousness" of Presidential inauguration speeches could be rated by several people using a 1-10 scale. A child's reaction to his or her mother's departure could be categorized as "securely attached" or "anxiously attached" based on certain well-specified behaviors. The essential feature of quantitative research is the existence of these codes, regardless of whether they are categorical (e.g., "securely attached") or numeric (e.g., 1-10 rating of ambitiousness), and the existence of a set of rules for using the codes.

Advantages. There are three major advantages of quantitative methods: reliability, falsifiability, and replication. The first advantage of transforming observations into codes is that it becomes possible to evaluate the **reliability** of the measure you are using. If the same child is consistently categorized as anxiously attached by different observers on different occasions, then we can have more confidence in our measure of attachment style. Without a coding system, there is no objective method of comparing the observations of multiple observers or of the same observer on multiple occasions. Thus, quantitative methods permit us to detect and disregard unreliable measures. The second advantage is that it is much easier to design a **falsifiable** study with quantitative methods. If our data have been coded as numbers, we might expect the average for two groups to differ: the average ambitiousness of wartime Presidents could be predicted to be higher than the average ambitiousness of peacetime Presidents. If our data have been coded as categories, we might expect a greater abundance of one category in one group than in another (e.g., a greater percentage of anxiously attached children in families with divorced parents than in families without divorced parents). Because it is easy to describe the pattern of results that would confirm or disconfirm these hypotheses before data are collected, the hypotheses are falsifiable. Finally, quantitative studies are easier to **replicate** because later researchers can adopt the explicit coding systems described by the original researchers.

Disadvantages. The major disadvantages to quantitative research are that 1) it presumes a considerable degree of understanding of the phenomenon of interest and 2) it may neglect aspects of the phenomenon that are unexpected. First, quantitative research presumes that the researcher understands a phenomenon well enough to reliably measure it. For new or rare phenomena, this may be unrealistic. For example, how do we code the visual hallucinations experienced by someone with a rare brain trauma? The second disadvantage is that the coding system you select may neglect some important aspect of the phenomenon. For example, you may be studying schoolyard bullying when you unexpectedly witness one child humiliate a child in front of their peers and then console that child when they are alone. It is very likely that your coding system has no code for this pattern, yet it could be very important.

Appropriate when: testing a hypothesis. Quantitative measures are appropriate for situations where you have sufficient control to reliably measure the phenomenon you are interested in and sufficient background knowledge to formulate a hypothesis.

Qualitative

Before contrasting quantitative research with qualitative research, it is worth noting that there is widespread disagreement in how the words "qualitative" and "quantitative" are used with regard to research. *Quantitative* is sometimes interpreted as "using numbers" while *qualitative* is "using categories" (e.g., anxious, secure). As discussed above, I prefer to group these together under the single heading "quantitative." The reason for this is that there is a more profound difference in how research is conducted that is not captured by the simple "numbers vs. categories" distinction. According to my definition, **qualitative** research involves observations that are transformed into records based on the observer's judgment. Some examples include: an anthropologist describing the folkways of a culture, a neurologist describing the symptoms of a rare disorder (an example of a *case study*), or an ethologist describing the activities of a chimpanzee troop.

Advantages. The major advantage of qualitative research is that it can provide a very rich description of a phenomenon, including not only detail of the phenomenon but also extensive detail about the context in which the phenomenon was observed. Whereas quantitative research tends to take a more

narrow focus to *answer a particular question*, qualitative research tends to take a broader perspective to *describe a particular phenomenon.* It is also possible that the description generated by qualitative research may capture rare events that would be unlikely to be considered in a quantitative coding system.

Disadvantages. The major disadvantages of qualitative research are the mirror images of the advantages of the quantitative approach: difficult to determine reliability, difficult to falsify, and difficult to replicate. Two anthropologists may differ markedly in their interpretation of the same culture, but with no coding system it is difficult to estimate the degree of (un)reliability in their assessments. The lack of coding system also makes replication difficult. Falsifiability is compromised because observers' intuitions may lead them to seek out, notice, interpret, and remember events that are consistent with their expectations – the **confirmation bias.** Although quantitative research is also vulnerable to the confirmation bias, its vulnerability is reduced by a more explicit separation of observations and interpretations.

Appropriate when: the goal is *description* or when a phenomenon is new or rare, such as the three examples cited earlier of the anthropologist, neurologist, and ethologist.

In naturalistic observation. Researchers using naturalistic observation frequently use qualitative methods to interpret their data. To reduce the risk of spurious interpretations, these researchers rely on **multiple confirmations** of a phenomenon and **negative case analysis** to explain anomalies. First, researchers must have *multiple confirmations* of a phenomenon (e.g., Moore's (1985) observation of women initiating flirting by flipping their hair) before they identify it as a phenomenon. Second, researchers who identify a consistent pattern but witness a "negative case" (an inconsistency in the expected pattern) are obligated to explain that inconsistency.

Basic and Applied Research

Applied research is directed toward the solution of a particular problem. A lot of research on mental illness is applied because it is directed toward the treatment or cure of mental illness. In contrast, **basic research** is directed toward understanding the fundamental nature of a phenomenon. For example, research on the role of neurotransmitters in attention could be considered basic research. Basic research is sometimes criticized because the usefulness of its results may not be immediately apparent. With time, however, basic research can often be applied to a particular problem. Schizophrenia involves both problems with neurotransmitter levels and with attention, so basic research on these issues could help to solve the problem of how to treat schizophrenia.